(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号
特開2000－66691
（P2000－66691A）
(43)公開日　平成12年３月３日(2000.3.3)

| (51)Int.Cl.⁷ | 識別記号 | ＦＩ | | テーマコード（参考） |
|---|---|---|---|---|
| Ｇ１０Ｌ | 11/00 | Ｇ１０Ｌ | 7/02 | Ａ |
| | 11/06 | | 9/00 | Ｃ |
| | 19/00 | | 9/16 | |
| | | | 9/18 | Ｈ |

審査請求　未請求　請求項の数12　ＯＬ　（全 15 頁）

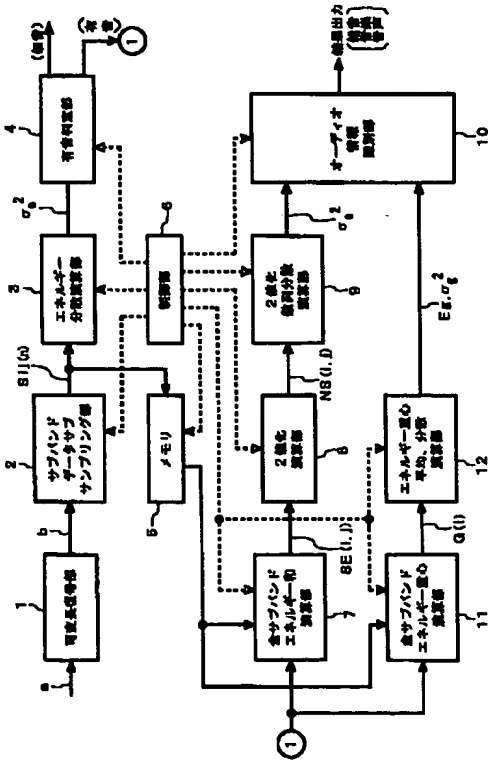
(21)出願番号　特願平10－235543
(22)出願日　平成10年８月21日(1998.8.21)

(71)出願人　000001214
ケイディディ株式会社
東京都新宿区西新宿２丁目３番２号
(72)発明者　中島　康之
東京都新宿区西新宿２丁目３番２号　国際電信電話株式会社内
(72)発明者　菅野　勝
東京都新宿区西新宿２丁目３番２号　国際電信電話株式会社内
(74)代理人　100084870
弁理士　田中　香樹　（外１名）

最終頁に続く

(54)【発明の名称】　オーディオ情報分類装置

(57)【要約】
【課題】　簡単かつ高速に、無音／有音区間の判別、音楽区間と音声区間、あるいは音楽区間と音声区間と雑音区間に分類することを可能とするオーディオ情報分類装置を提供することにある。
【解決手段】　有音判定部４はエネルギー分散演算部３で求められた値$\sigma_e^2$が閾値より大きい時有音と判定する。有音と判定されると、メモリ５に格納されていたオーディオ情報が読み出されて、全サブバンドエネルギー和演算部７とサブバンドエネルギー重心演算部１１に入力される。前記演算部７の出力は２値化演算部８で２値化され、２値化数列分散演算部９で２値化数列の単位時間内の分散$\sigma_s^2$が求められる。一方、エネルギー重心平均・分散演算部１２はエネルギー重心平均Ｅｇと分散$\sigma_g^2$を求める。オーディオ情報識別部１０は、前記分散$\sigma_s^2$、エネルギー重心平均Ｅｇ、および分散$\sigma_g^2$に対して識別関数を用いて、雑音、音楽、音声の判別を行う。



【特許請求の範囲】

【請求項１】　オーディオ情報から音声区間と音楽区間を分類するオーディオ情報分類装置において、
入力されたオーディオ情報から単位時間ごとの周波数データを抽出するオーディオ周波数データ抽出手段と、
抽出した単位時間ごとの周波数データのエネルギーの分散を求め、分散値の大きさにより無音／有音区間を判定する無音／有音判定手段とを具備することを特徴とするオーディオ情報分類装置。

【請求項２】　請求項１に記載のオーディオ情報分類装置において、
前記オーディオ周波数データ抽出手段によって抽出される単位時間ごとの周波数データは、入力されたオーディオ情報がMPEGデータである場合、単位時間分のMPEG符号化データにおける最低周波数成分のエネルギーの分散を利用することを特徴とするオーディオ情報分類装置。

【請求項３】　オーディオ情報から音声区間と音楽区間を分類するオーディオ情報分類装置において、
入力されたオーディオ情報から有音部のみを抽出する有音抽出手段と有音区間における音の疎密度により音声であるか音楽であるかを判定する音声／音楽区間判定手段とを具備することを特徴とするオーディオ情報分類装置。

【請求項４】　請求項３に記載のオーディオ情報分類装置において、
疎密度はオーディオ信号のエネルギーの大きさによって２値化された数列の分散を用いて疎密度を判定することを特徴とするオーディオ情報分類装置。

【請求項５】　請求項４に記載のオーディオ情報分類装置において、
前記オーディオ信号のエネルギーは、入力されたオーディオ情報がMPEGデータである場合、単位時間分のMPEG符号化データにおける全周波数成分のエネルギー和を利用することを特徴とするオーディオ情報分類装置。

【請求項６】　請求項３ないし請求項５のいずれかに記載のオーディオ情報分類装置において、
前記音声／音楽区間判定手段は疎密度を特徴ベクトルとしたBayes 決定則を用いて、テストデータに対して音楽と音声区間の共分散行列を求めておき、入力データに対して正規分布パターンにおけるBayes 決定識別関数を用いて各音楽区間と音声区間の判別を行うことを特徴とするオーディオ情報分類装置。

【請求項７】　オーディオ情報から音声区間と音楽区間を分類するオーディオ情報分類装置において、
入力されたオーディオ情報から有音部のみを抽出する有音抽出手段と入力されたオーディオ情報から有音時の単位時間ごとの周波数データを抽出するオーディオ周波数データ抽出手段と、
オーディオ周波数データから単位時間における周波数の重心の平均と重心の標準偏差を求め、周波数の重心の分布により雑音区間か否かを判別する雑音区間抽出手段を具備することを特徴とするオーディオ情報分類装置。

【請求項８】　請求項７に記載のオーディオ情報分類装置において、
前記オーディオ周波数データ抽出手段によって抽出される単位時間ごとの周波数データは、入力されたオーディオ情報がMPEGデータである場合、単位時間分のMPEG符号化データにおける周波数成分のエネルギーの重心を利用することを特徴とするオーディオ情報分類装置。

【請求項９】　請求項７又は８に記載のオーディオ情報分類装置において、
前記雑音抽出手段は、周波数成分の重心の平均と分散を特徴ベクトルとしたBayes 決定則を用いて、テストデータに対して雑音と雑音以外の共分散行列を求めておき、入力データに対して正規分布パターンにおけるBayes 決定識別関数を用いて各雑音区間と非雑音区間の判別を行うことを特徴とするオーディオ情報分類装置。

【請求項１０】　オーディオ情報から音声区間と音楽区間を分類するオーディオ情報分類装置において、
入力されたオーディオ情報から有音部のみを抽出する有音抽出手段と、
入力されたオーディオ情報から有音時の単位時間ごとの周波数データを抽出するオーディオ周波数データ抽出手段と、
オーディオ周波数データの単位時間における疎密度および単位時間における周波数の重心の平均と重心の標準偏差を特徴ベクトルとしたBayes 決定則を用いて、テストデータに対して音声と音楽と雑音の共分散行列を求めておき、入力データに対して正規分布パターンにおけるBayes 決定識別関数を用いて音声、音楽、雑音区間の判別を行うことを特徴とする音声／音楽／雑音区間判別手段を具備することを特徴とするオーディオ情報分類装置。

【請求項１１】　オーディオ情報から音声区間と音楽区間を分類するオーディオ情報分類装置において、
入力されたオーディオ情報から有音部のみを抽出する有音抽出手段と、
入力されたオーディオ情報から有音時の単位時間ごとの周波数データを抽出するオーディオ周波数データ抽出手段と、
オーディオ周波数データの単位時間における周波数の重心の平均と重心の標準偏差を特徴ベクトルとしたBayes 決定則を用いて、テストデータに対して雑音と雑音以外の共分散行列を求めておき、入力データに対して正規分布パターンにおけるBayes 決定識別関数を用いて雑音と雑音以外の区間の判別を行うことを特徴とする雑音区間判別手段と、
オーディオ周波数データの単位時間における疎密度を特徴ベクトルとしたBayes 決定則を用いて、テストデータに対して音声と音楽の共分散行列を求めておき、前記、雑音区間判別手段で雑音以外と判別された区間に対し

て、正規分布パターンにおけるBayes 決定識別関数を用いて音声、音楽、雑音区間の判別を行うことを特徴とする音声／音楽区間判別手段を具備するオーディオ情報分類装置。

【請求項１２】　請求項３ないし請求項１１のいずれかに記載のオーディオ情報分類装置において、
入力されたオーディオ情報から有音部のみを抽出する有音抽出手段は、請求項１または２に示された有音判定手段を用いることを特徴とするオーディオ情報分類装置。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明はオーディオ情報の分類装置に関し、特に符号化されていない元のままのオーディオ情報あるいは符号化されたオーディオ情報のいずれからも、簡単かつ高速に、音声区間と音楽区間、あるいは音声区間と音楽区間と雑音区間を分類できるオーディオ情報の分類装置に関する。

【０００２】

【従来の技術】インターネットに代表されるように、分散したデータベースに、テキストのみならず音声や映像情報が蓄積される技術分野においては、マルチメディア情報を効果的にインデックスする方法が必要とされている。このうちオーディオ信号を分類する手法については、オーディオ信号を音楽や音声区間に分類することで、おおまかなインデックスが可能になる。例えば、E. ScheirerとM. Slaneyの"Construction and evaluation of a robust multifeature speech/music discriminator, Proceedings of IEEE ICASSP, pp.1331-1334, 1997ではオーディオ信号について４Hz 成分、フレーム間スペクトル差分、パルス検出の３つの特徴パラメータを利用してBayes 決定法などの識別関数により音声と音楽の判別を行っている。

【０００３】図１３は前記分類を行う手法の説明図である。オーディオ信号Ａは４Ｈｚ帯域フィルタ２１、周波数変換部２２、およびサブバンド分割部２３に入力する。４Ｈｚ帯域フィルタ２１はオーディオ信号Ａの４Ｈｚ成分を抽出し、４Ｈｚ帯域エネルギ演算部２４に出力する。周波数変換部２２はオーディオ信号Ａをスペクトル分析し、フレーム間スペクトル差分演算部２５に出力する。また、サブバンド分割部２３でサブバンド分割されたオーディオ信号Ａは、包絡線ピーク検出部２６に出力される。

【０００４】一般に、４Ｈｚ成分については、音声信号ではこの周波数成分が特に強く出現する特徴がある。フレーム間のスペクトル差分については、音楽のように変化の激しい場合に大きくなる特徴がある。さらに，パルス検出は入力信号を各周波数帯域（サブバンド）に分け包絡線のピークを検出する。音楽のようにリズムのあるオーディオ信号では全ての帯域において周期的にこのピークが現れる。

【０００５】オーディオ情報識別部２７は、前記の４Ｈｚ成分、フレーム間スペクトル差分、パルス検出の３つの特徴パラメータを利用して、Bayes 決定法などの識別関数により音声と音楽の判別を行う。なお、入力してくるオーディオ信号が圧縮符号化されたオーディオ信号である場合には、図示されていない復号処理部で復号して、前記４Ｈｚ帯域フィルタ２１、周波数変換部２２およびサブバンド分割部２３に送出する。

【０００６】

【発明が解決しようとする課題】しかしながら、前記の従来技術は、圧縮符号化されたオーディオ信号から音楽区間、音声区間を検出する場合には、一旦圧縮されたデータを復号してアナログのオーディオ信号Ａに戻してから検出処理を行うことになり、処理時間も大幅に増加するという問題点がある。

【０００７】また、雑音成分の除去についてはなんら触れておらず、雑音成分を除いた音声や音楽区間の判別ができないという問題がある。換言すれば、雑音成分も音声や音楽区間に含められてしまうという問題がある。

【０００８】本発明の目的は、前記した従来技術の問題点に鑑み、符号化されていない元のままのオーディオ情報あるいは圧縮符号化されたオーディオ情報のいずれからも、簡単かつ高速に、無音／有音区間の判別、音楽区間と音声区間、あるいは音楽区間と音声区間と雑音区間に分類することを可能とするオーディオ情報分類装置を提供することにある。他の目的は、雑音区間を除去して、音楽区間、音声区間を検出することが可能なオーディオ情報分類装置を提供することにある。

【０００９】

【課題を解決するための手段】前記した目的を達成するために、本発明は、入力されたオーディオ情報から単位時間ごとの周波数データを抽出するオーディオ周波数データ抽出手段と、抽出した単位時間ごとの周波数データのエネルギーの分散を求め、分散値の大きさにより無音／有音区間を判定する無音／有音判定手段とを具備した点に第１の特徴がある。

【００１０】また、入力されたオーディオ情報から有音部のみを抽出する有音抽出手段と、有音区間における音の疎密度により音声であるか音楽であるかを判定する音声／音楽区間判定手段とを具備した点に第２の特徴がある。

【００１１】また、入力されたオーディオ情報から有音部のみを抽出する有音抽出手段と、入力されたオーディオ情報から単位時間ごとの周波数データを抽出するオーディオ周波数データ抽出手段と、オーディオ周波数データから単位時間における周波数の重心の平均と重心の標準偏差を求め、周波数の重心の分布により雑音区間か否かを判別する雑音区間抽出手段を具備した点に第３の特徴がある。

【００１２】さらに、オーディオ周波数データの単位時

間における疎密度および単位時間における周波数の重心の平均と重心の標準偏差を特徴ベクトルとしたBayes 決定則を用いて、テストデータに対して音声と音楽と雑音の共分散行列を求めておき、入力データに対して正規分布パターンにおけるBayes 決定識別関数を用いて音声、音楽、雑音区間の判別を行う音声／音楽／雑音区間判別手段を具備した点に第４の特徴がある。

【００１３】さらに、オーディオ周波数データの単位時間における周波数の重心の平均と重心の標準偏差を特徴ベクトルとしたBayes 決定則を用いて、テストデータに対して雑音と雑音以外の共分散行列を求めておき、入力データに対して正規分布パターンにおけるBayes 決定識別関数を用いて雑音と雑音以外の区間の判別を行うことを特徴とする雑音区間判別手段と、オーディオ周波数データの単位時間における疎密度を特徴ベクトルとしたBayes 決定則を用いて、テストデータに対して音声と音楽の共分散行列を求めておき、前記、雑音区間判別手段で雑音以外と判別された区間に対して、正規分布パターンにおけるBayes 決定識別関数を用いて音声、音楽、雑音区間の判別を行うことを特徴とする音声／音楽区間判別手段を具備した点に第５の特徴がある。

【００１４】本発明によれば、符号化されていないもとのままのオーディオ情報、あるいは符号化されたオーディオ情報のいずれからも、簡単かつ高速に、無音／有音区間の判別、あるいは音声区間、音楽区間、雑音区間を分類することが可能になる。

【００１５】

【発明の実施の形態】以下に、図面を参照して、本発明を詳細に説明する。この実施形態は動画像および音声符号化の国際標準であるMPEG１（ISO/IEC 11172 ）およびMPEG2(ISO/IEC 13818)により圧縮されたオーディオ符号化データを用いて音声、音楽、雑音区間を分類するものであるが、本発明はこれに限定されるものではない。図１は本発明のオーディオ情報分類装置の一実施形態のブロック図を示す。また、図２は本実施形態の動作を説明するフローチャートである。

【００１６】図１に示されているように、圧縮符号化されたオーディオ符号化データａは可変長復号部１に入力される。ここで、圧縮符号化されたオーディオの符号化データ構造について、MPEG1 レイヤーIIを例にして図４を参照して説明する。MPEG1 では図示されているように、元のオーディオ信号ｐからサンプリングした５１２個のＰＣＭサンプルをサブバンド符号化して、３２個のサブバンドデータＰi(n)（n=0,1,...,31）を作り、それを時間的にサンプルをずらしながら３６回(i=0,1,...,35)繰り返して合計１１５２個のサブバンドデータを作り、この１１５２個のサブバンドデータを１フレームの符号化データＱとしている。

【００１７】前記した構造の符号化データＱが前記可変長復号部１に連続して入力してくると、該可変長復号部１にはこれを各フレームのサブバンドデータに復号し、サブバンドデータサンプリング部２に出力する。いま、ある単位時間を１秒とすると、この１秒は図５のａのように３８フレームから構成されているので、可変長復号部１は１秒分の符号化データに対し、同図のｂのように３８個の３２サブバンド×３６サンプルを出力する。

【００１８】サブバンドデータサンプリング部２では、図５のｃに示されているように、単位時間（例えば１秒）分のサブバンドデータのうち、各フレームｉのｊ番目(j=0,1,....,35 は１フレーム内のサンプル数）にあるサブバンドデータＳij(n)(i=0,1,....,37は単位時間内のフレーム数）を抽出し、図１のエネルギー分散演算部３およびメモリ５に入力する。該サブバンドデータサンプリング部２は、入力されたオーディオ情報から単位時間ごとの周波数データを抽出するオーディオ周波数データ抽出手段と呼ぶことができる。

【００１９】以上の動作は、図２では、ステップＳ１～Ｓ９で行われる。ステップＳ１では、フレーム番号を表すｉが０と置かれ、ステップＳ２ではサブバンド番号を表すｎが０と置かれる。ステップＳ３では、可変長復号部１にて符号化データが可変長復号され、ステップＳ４ではｉフレーム目のｊサンプル目のサブバンドデータＳi,j(n)が抽出される。次に、ステップＳ５にてｎ＝３２が成立するか否かの判断がなされ、この判断が否定のときはステップＳ６に進んでｎに１が加算される。そしてステップＳ３に戻って前記と同様の処理が行われる。以上のステップＳ３～Ｓ６の処理が繰り返して行われて、ステップＳ５の判定が肯定となると、サブバンドデータサンプリング部２から、フレームｉ、サンプルｊのサブバンドデータＳi,j(n)が抽出されたことになる。

【００２０】ステップＳ５の判断が肯定になるとステップＳ７に進み、ｉに１が加算される。次にステップ８に進み、ｉ＝Ｎf が成立するか否かの判断がなされる。ここで、Ｎf は単位時間内のフレーム数である。この判断が否定の場合はステップＳ２に戻り、再びｎ＝０とされて、再度前記した処理が行われる。以上の処理が繰り返し行われ、ステップＳ８の判断が肯定になると、ｉ＝０～（Ｎf －１）フレームの各ｊ番目のサンプルのサブバンドデータＳi,j(n)が抽出されたことになり、ステップＳ９にてこれらのサブバンドデータＳi,j(n)は図１の各フレームのエネルギー分散演算部３およびメモリ５へ転送される。

【００２１】エネルギー分散演算部３では、図６の（１）および（２）式に従って、単位時間当たりのエネルギー分散$\sigma_e^2$を計算し、有音判定部４に入力する。なお、（１）式で、Ｎf は単位時間内のフレーム数、Ｎj は１フレーム中のサンプル数で、例えばＮj を１とした場合、フレーム中の先頭のサンプルのみを用いて計算することになり、処理の高速化を図ることが可能である。また、サブサンプルデータＳi,j(n)でｎ＝０と

すると、低周波成分のみを用いてエネルギー分散$\sigma_e^2$を計算することになり、この場合、高周波成分までを含んだ場合と同等な結果が得られ、処理時間も高速化することが可能である。

【００２２】有音判定部４では、入力された単位時間における音声情報が無音であるか有音であるかを下記の（３）式にしたがって判定し、条件に合う場合は有音であると判定する（ステップＳ１１）。有音である場合は、無音である場合に比べて、単位時間のエネルギー分散が大きいから下記の（３）式が成立することになる。

$$\sigma_e^2 > \alpha \qquad (3)$$

ここに、$\alpha$は予め定められた第１の閾値である。

【００２３】該有音判定部４において、入力された単位時間のオーディオ情報が有音であると判断された場合には、メモリ５から該単位時間内の周波数データすなわちサブバンドデータSi,j(n)を読み出して、全サブバンドエネルギー和演算部７（図３のステップＳ１２）とサブバンドエネルギー重心演算部１１（ステップＳ１６）に入力する。この機能は、オーディオ周波数データ抽出手段と呼ぶことができる。一方、無音であると判定された場合には、以降のオーディオ情報判定処理を終了し、ステップＳ１に戻る。

【００２４】全サブバンドエネルギー和演算部７では、図６の（４）式に従って、全サブバンドのエネルギー和SE(i,j)を計算し、２値化演算部８（ステップ１３）に入力する。SE(i,j)は３２バンド分のSi,j(n)のエネルギーの累積和である。２値化演算部８では、図６の（５）式に従って、Th1を基にSE(i,j)を２値化して、数列NS(i,j)を計算する。Th1はあらかじめ定められた２値化のための閾値である。

【００２５】音声と音楽の波形は図８のように、音声では断続した波形を持つのに対して、音楽では連続的な波形となる。これらの波形を２値化（正規化）すると、図８の右側の図から明らかなように、音の断続性がより明確になる。すなわち、有音区間における音の疎密度により音声であるか音楽であるかを判定できる。

【００２６】２値化演算部８で得られた２値化数列NS(i,j)は２値化数列分散演算部９（図３のステップＳ１４）に入力する。２値化数列分散演算部９では、２値化数列の単位時間内の分散$\sigma_s^2$を、図６の（６）式に従って計算し、オーディオ情報識別部１０に入力する（ステップＳ１５）。$\sigma_s^2$はNS(i,j)が０となるサンプル数の分散で、音声区間では断続性が強いため、該分散値は音楽区間に比べて大きくなる。この分散は、音の疎密度を表している。

【００２７】図６の（６）式で、MはNS(i,j)が単位時間内に１から０に変化する数で、単位時間内の０連続区間の個数を表す。また、Nns(k)はNS(i,j)が０の場合の連続数で、音楽のようにリズムがある場合は時間的な変化は小さい。

【００２８】サブバンドエネルギー重心演算部１１（ステップＳ１６）では、図７の（７）式に従って、フレームｉにおけるサブバンド重心G(i)が計算され、エネルギー重心平均、分散演算部１２（ステップＳ１７）に入力する。（７）式で、サブバンドの重心はすべてのサブバンドｎについて、各フレーム内のサンプルｊについて計算されるが、エネルギー分散$\sigma_e^2$の場合と同様に、Nj＝１としても重心値に大きな変化がなく、すべてのサンプルについて計算する場合よりも処理時間を削減することが可能である。

【００２９】エネルギー重心平均、分散演算部１２では、図７の（８）式および（９）式に従って単位時間内の分散$\sigma_g^2$とエネルギー重心の平均Egが計算され、オーディオ情報識別部１０（ステップＳ１８）に入力する。図９は単位時間を１秒としたときのサブバンドエネルギー重心の平均と分散の分布例であるが、歓声などの雑音は、音楽や音声などの他の音源と異なって、ある一定の領域ａに集中している。

【００３０】オーディオ情報識別部１０では、入力された２値化数列分散$\sigma_s^2$、サブバンドエネルギー重心平均Egおよび分散$\sigma_g^2$に対して、既知のBayes決定ルールに基づいた正規分布の場合の識別関数（図７の（１０）式）を用いて、雑音、音楽、音声の判別が行われる。ここで、クラスは雑音、音楽、音声の３つのクラスに分類する。また、入力ベクトルｘは（$\sigma_s^2$,Eg,$\sigma_g^2$）の要素で構成される。なお、（１０）式におけるmk,ck,p($\omega$k)は、トレーニングデータを用いて、あらかじめ求めておくことができる。判定は、入力ベクトルに対して、最も大きなfk(x)を与えるクラスｋが求める判別クラスとなり、結果を出力する。すなわち、トレーニングにより予め求められた各クラス（雑音、音楽、音声）のデータmk,ck,p($\omega$k)を（１０）式に代入し、これに前記（６）（９）（８）式で求められた入力ベクトルｘ（$\sigma_s^2$,Eg,$\sigma_g^2$）を入れて、各クラスの識別値fk(x)を求める。そして、該識別値fk(x)の一番大きいクラスが雑音であれば雑音、音声であれば音声、音楽であれば音楽と判定する。なお、オーディオ情報識別部１０は、Ｋ近傍決定則，ゆう度検定，Ｋ－平均法，Ｋ－決定木法などのような前記（１０）式以外の他の式を用いてクラスの判別をするようにしても良い。

【００３１】次に、本発明の第２の実施形態について、図１０を参照して説明する。図１０において、図１と同一または同等物には同じ符号が付されている。図１０の可変長復号部１～有音判定部４の動作（図２のステップＳ１～Ｓ１１）は前記第１実施形態と同じであるので、説明を省略し、サブバンドエネルギー重心演算部１１以降の動作を、図１１を参照して説明する。

【００３２】有音判定部４において、入力された単位時間のオーディオ情報が有音であると判断された場合に

は、メモリ５から単位時間内のサブバンドデータＳi,j(n)を読み出してサブバンドエネルギー重心演算部１１に入力する。一方、無音であると判定された場合には、以降のオーディオ情報判定処理を終了し、ステップＳ１に戻る。

【００３３】サブバンドエネルギー重心演算部１１（ステップＳ１６）では、図７の（７）式に従って、フレームｉにおけるサブバンド重心Ｇ(i)が計算され、エネルギー重心平均、分散演算部１２（ステップＳ１７）に入力する。（７）式で、サブバンドの重心は全てのサブバンドｎについて、各フレーム内のサンプルｊについて計算されるが、エネルギー分散$\sigma_e^2$の場合と同様に、Ｎj＝１としても重心値に大きく変化がなく、すべてのサンプルについて計算する場合よりも処理時間を削減することが可能である。

【００３４】エネルギー重心平均、分散演算部１２では（８）式および（９）式に従って単位時間内の分散$\sigma_g^2$とエネルギー重心の平均Ｅg が計算され、雑音識別部１３（ステップＳ１８）に入力する。

【００３５】雑音識別部１３では、入力されたサブバンドエネルギー重心平均Egおよび分散$\sigma_g^2$に対してBayes 決定ルールに基づいた正規分布の場合の識別関数（１０）式を用いて、雑音か否かの判別が行われる。ここで、クラスは雑音と雑音外の２つに分類する。また、入力ベクトルｘは（Eg, $\sigma_g^2$）の要素で構成される。（１０）式におけるmk,ck,p(ωk)は、トレーニングデータを用いて予め求めておくことができる。判定は、入力ベクトルに対して、最も大きなｆk(x)を与えるクラスｋが求める判別クラスとなり、結果を出力する。

【００３６】ここで、雑音と判定された場合（ステップＳ３０が肯定）は、雑音である旨の結果を出力後、最終データでない限り（ステップＳ２３が否定）、次のデータ入力を行う。また、雑音外と判定された場合（ステップＳ３０が否定）は、次の処理（ステップ１２）へ進み、音楽か音声の判定を行う。

【００３７】音楽か音声の判定処理に進むと、メモリ５から全サブバンドエネルギー和演算部７にＳi,j(n)が入力され、全サブバンドエネルギー和演算部７では、図６の（４）式に従って、全サブバンドのエネルギー和ＳＥ(i,j)を計算し、２値化演算部８（ステップ１３）に入力する。ＳＥ(i,j)は３２バンド分のＳi,j(n)のエネルギーの累積和である。２値化演算部８では、図６の（５）式に従って、ＳＥ(i,j)を２値化して、数列ＮＳ(i,j)を計算する。Thlは予め定められた２値化のための閾値である。

【００３８】２値化演算部８で得られた２値化数列ＮＳ(i,j)は２値化数列分散演算部９（ステップ１４）に入力する。２値化数列分散演算部９では、２値化数列の単位時間内の分散$\sigma_s^2$を図６の（６）式にしたがって計算し、音楽音声識別部１４に入力する（ステップ１５）。$\sigma_s^2$はＮＳ(i,j)が０となるサンプル数の分散で、音声区間では断続性が強いため、該分散値は音楽区間に比べて大きくなる。

【００３９】音楽音声識別部１４では、入力された２値化数列分散$\sigma_s^2$に対してBayes 決定ルールに基づいた正規分布の場合の識別関数（１０）式を用いて、音楽、音声の判別が行われる。ここで、クラスは音楽、音声の２つのクラスに分類する。また、入力ベクトルｘは（$\sigma_s^2$）の要素で構成される。さらに、（１０）式におけるmk,ck,p(ωk)は、トレーニングデータを用いて、予め求めておくことができる。判定は、入力ベクトルに対して、最も大きなｆk(x)を与えるクラスｋが求める判別クラスとなり、結果を出力する。

【００４０】以上のように、前記第１、第２実施形態によれば、圧縮符号化されたオーディオの符号化データから無音／有音を判別し、有音の場合、音楽区間、音声区間、雑音区間を区別し、それぞれのタイムコードを図示されていない音声区間保持部、音楽区間保持部、雑音区間保持部のそれぞれに記録させることができる。

【００４１】さらに、本発明は圧縮されていないオーディオ情報の分類に関しても適用できる。その場合の実施形態を以下に説明する。

【００４２】圧縮符号化されていないオーディオ情報を扱う場合は、図１の可変長復号部１およびサブバンドデータサブサンプリング部２は高速フーリエ変換部（以下ＦＦＴ変換部）に置き換えられる。元のオーディオ情報からこのＦＦＴ変換部において、図１２にあるようなＦＦＴ変換を行い、単位時間分の周波数データを抽出する。今、該単位時間を１秒とすると、元のオーディオ信号ｐからサンプリングした２０４８個のサンプルをＦＦＴ変換し、それを時間的にサンプルをずらしながら３８回繰り返して合計２０４８×３８個のＦＦＴデータを単位時間分の周波数データとしている。

【００４３】その後、各フレームのエネルギー分散、エネルギー重心演算の平均および分散、エネルギー和の２値化後の数列分散を計算して、無音／有音、音楽、音声、雑音の判定を行う。

【００４４】

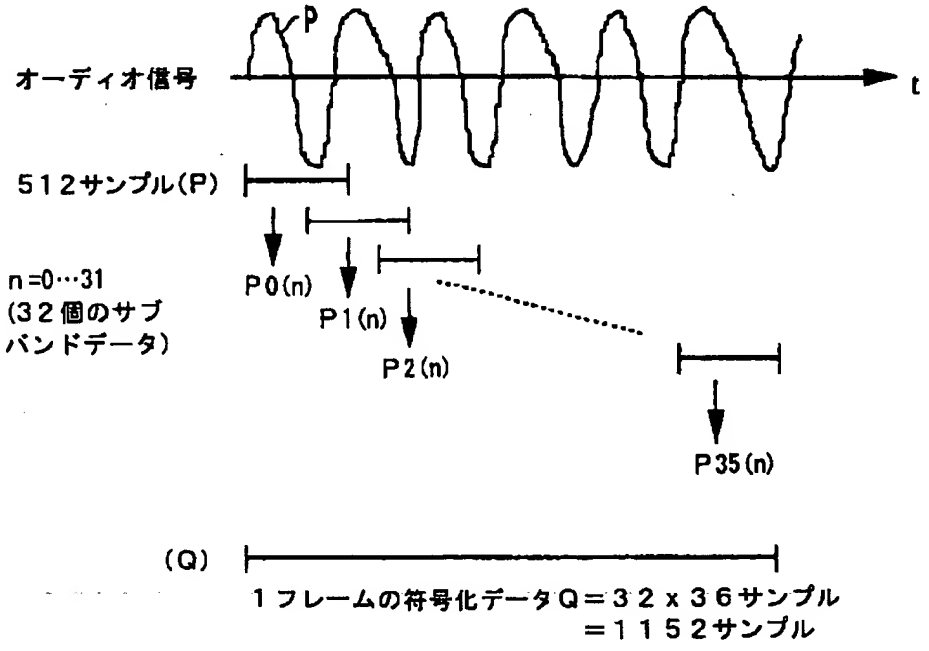
【発明の効果】以上の説明から明らかなように、本発明によれば、圧縮符号化されたあるいは圧縮符号化されていないオーディオデータから、符号化データ上で、オーディオ情報を有音／無音、音楽／音声／雑音区間に分類することが可能である。

【００４５】本発明を実際に動作させ、MPEG1 レイヤIIで符号化された１５分間のテレビ番組を用いて１秒毎の分類を行ったところ、無音の判定は９２％、音声区間の検出は９９％、音楽区間は７５％、雑音区間は７４％程度検出することが可能になった。

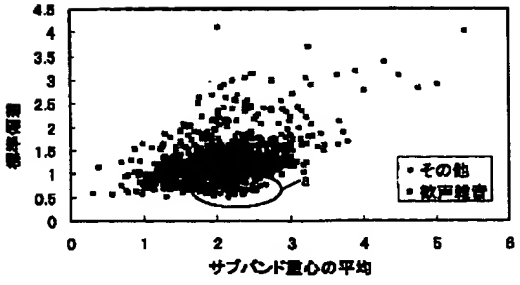
【図面の簡単な説明】

【図１】　本発明の一実施形態の構成を示すブロック図である。
【図２】　本実施形態の動作を示すフローチャートである。
【図３】　図２の続きのフローチャートである。
【図４】　MPEGオーディオ符号化データの構造を説明するための図である。
【図５】　図１のザブバンドデータサブサンプリング部の動作を説明するための図である。
【図６】　本実施形態で使用される数式を表す図である。
【図７】　本実施形態で使用される数式を表す図である。
【図８】　音声および音楽の正規化前および正規化後の波形図である。
【図９】　雑音のサブバンド重心の平均を表す図である。
【図１０】　本発明の第２実施形態の構成を示すブロック図である。
【図１１】　第２実施形態の要部の動作を示すフローチャートである。
【図１２】　符号化されていないオーディオ情報の周波数データの抽出方法を説明するための図である。
【図１３】　従来のオーディオ情報分類装置の構成を示すブロック図である。
【符号の説明】
1…可変長復号部、２…ザブバンドデータサブサンプリング部、３…エネルギー分散演算部、４…有音判定部、５…メモリ、６…制御部、７…全サブバンドエネルギー和演算部、８…２値化演算部、９…２値化数列分散演算部、１０…オーディオ情報識別部、１１…サブバンドエネルギー重心演算部、１２…エネルギー重心平均・分散演算部、１３…雑音識別部、１４…音楽音声識別部。

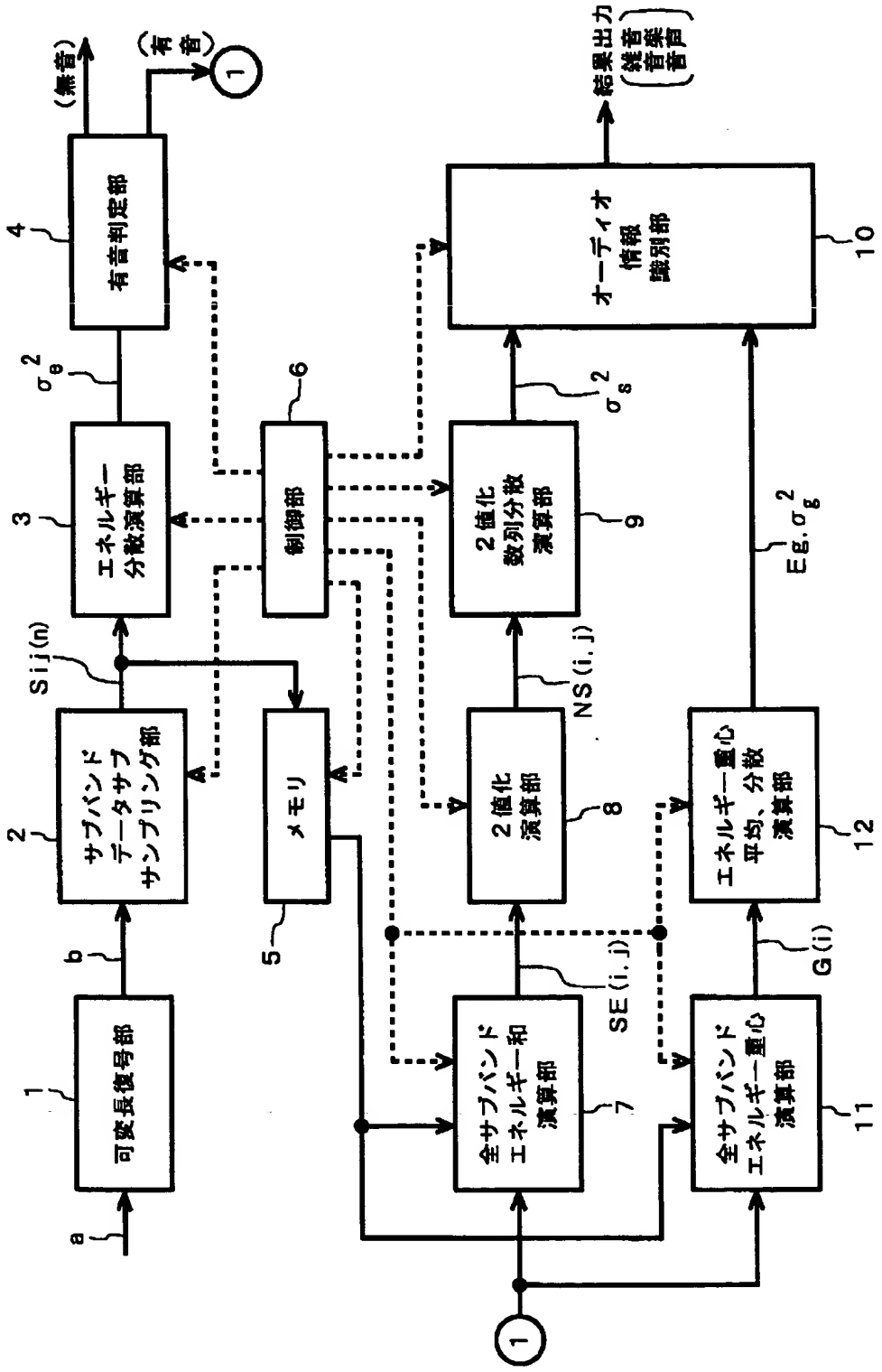
【図４】



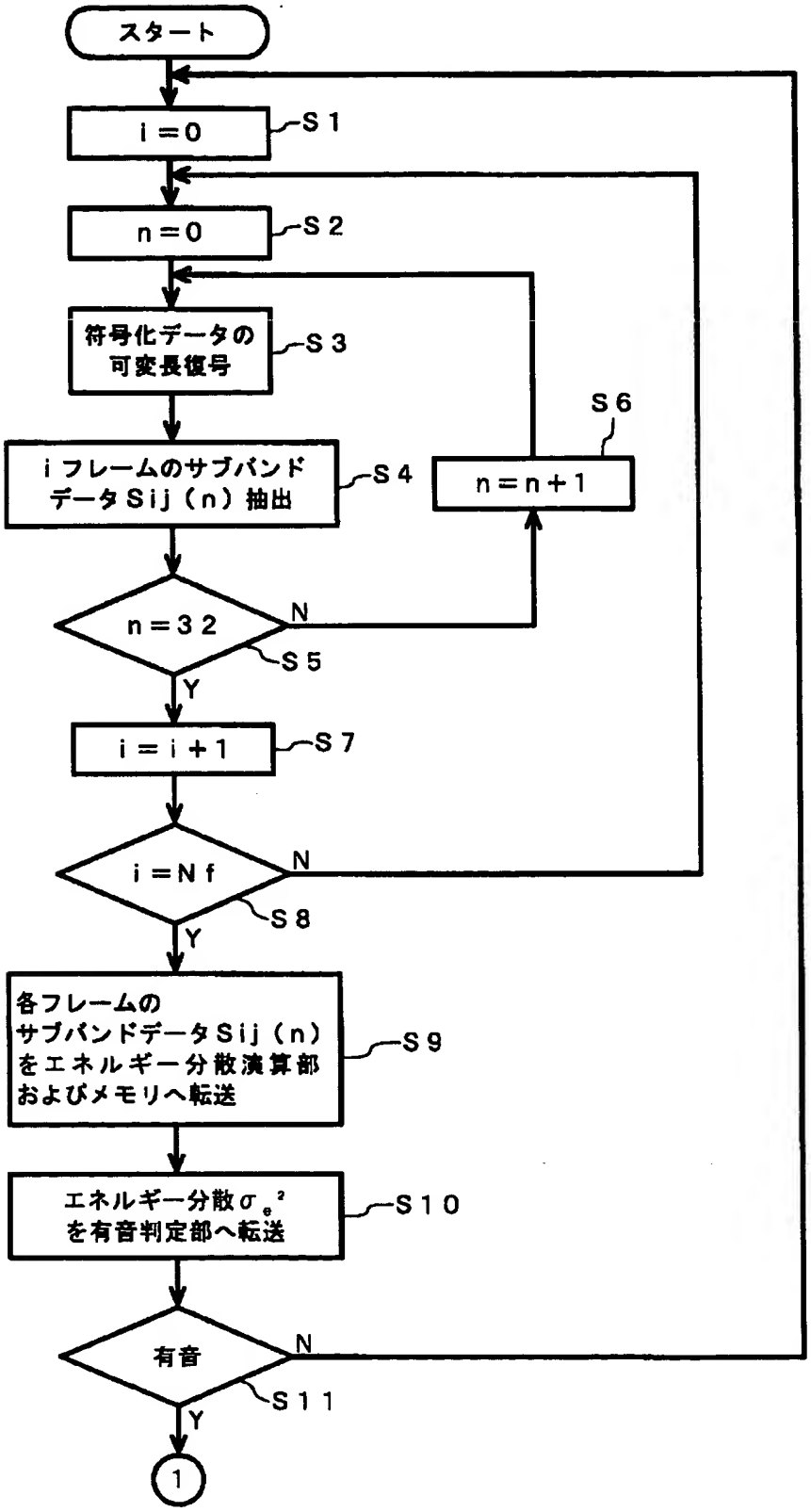

【図９】



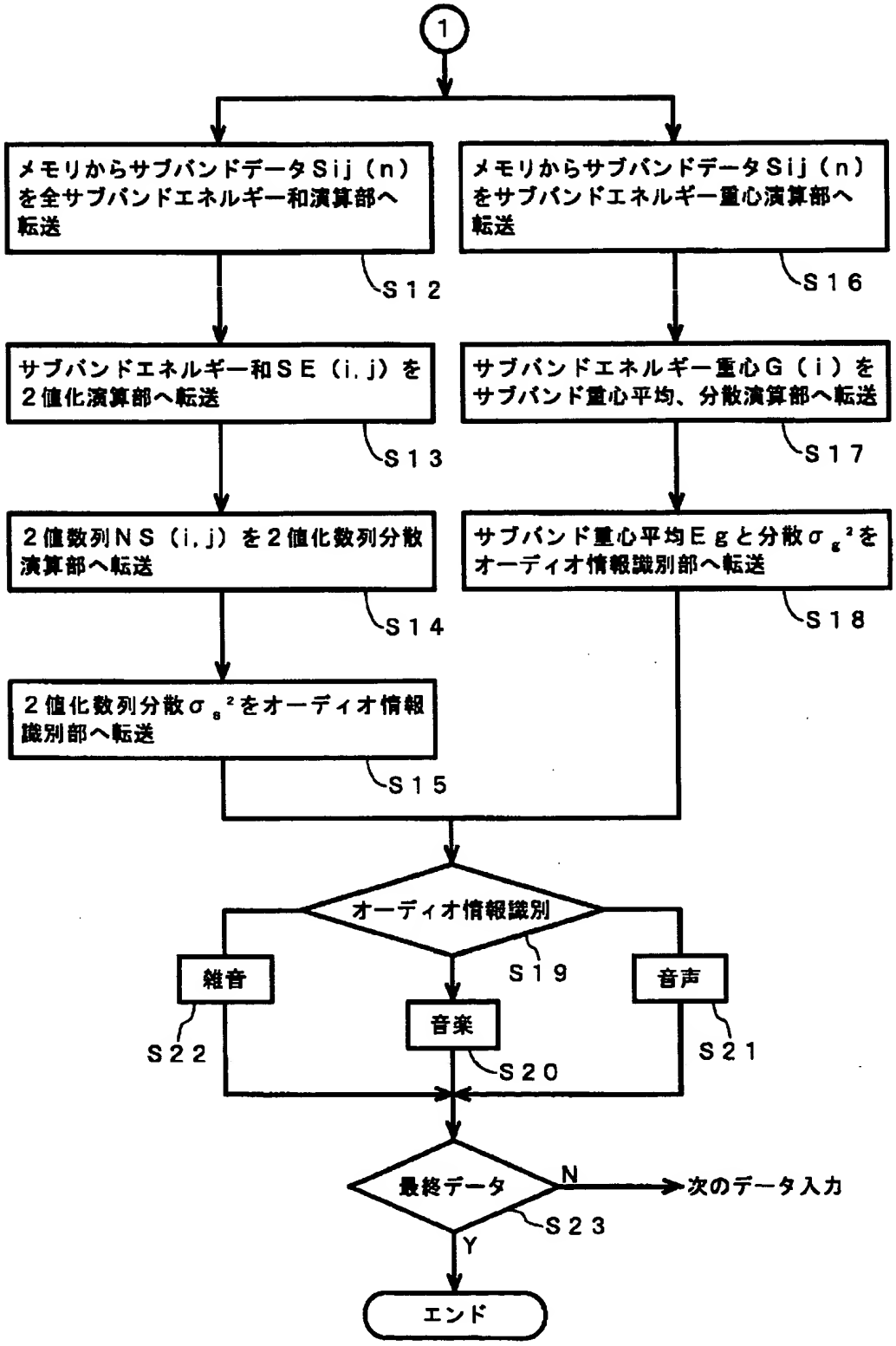
【図1】


可変長復号部
サブバンド データサブ サンプリング部
エネルギー 分散演算部
有音判定部
メモリ
制御部
全サブバンド エネルギー和 演算部
2値化 演算部
2値化 数列分散 演算部
全サブバンド エネルギー重心 演算部
エネルギー重心 平均、分散 演算部
オーディオ 情報 識別部
結果出力
(無音)
(有音)
a
b
Sij(n)
NS(i, j)
SE(i, j)
G(i)
Eg, σg2
σe2
σs2

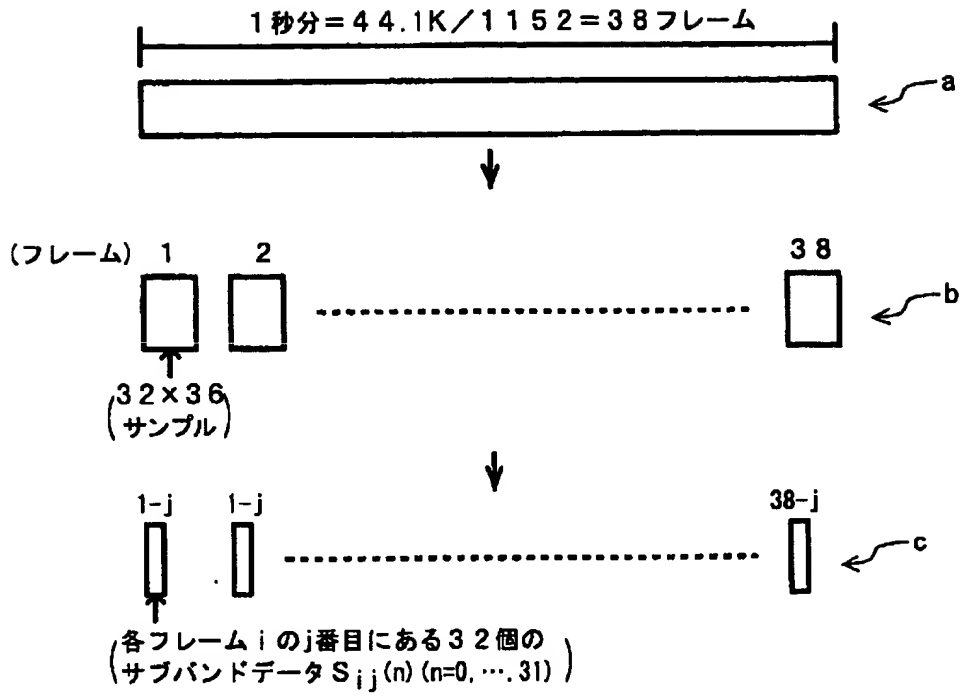
【図２】


スタート
i＝0
S１
n＝0
S２
符号化データの
可変長復号
S３
iフレームのサブバンド
データSij（n）抽出
S４
n＝n＋1
S６
n＝３２
N
S５
Y
i＝i＋1
S７
i＝Nf
N
S８
Y
各フレームの
サブバンドデータSij（n）
をエネルギー分散演算部
およびメモリへ転送
S９
エネルギー分散$\sigma_e^2$
を有音判定部へ転送
S１０
有音
N
S１１
Y
1

【図３】


①
メモリからサブバンドデータSij（n）を全サブバンドエネルギー和演算部へ転送
S12
メモリからサブバンドデータSij（n）をサブバンドエネルギー重心演算部へ転送
S16
サブバンドエネルギー和SE（i, j）を２値化演算部へ転送
S13
サブバンドエネルギー重心G（i）をサブバンド重心平均、分散演算部へ転送
S17
２値数列NS（i, j）を２値化数列分散演算部へ転送
S14
サブバンド重心平均Egと分散σg²をオーディオ情報識別部へ転送
S18
２値化数列分散σs²をオーディオ情報識別部へ転送
S15
オーディオ情報識別
S19
雑音
S22
音楽
S20
音声
S21
最終データ
N
次のデータ入力
S23
Y
エンド

【図５】




【図７】

$$G(i)=\sum_{n=0}^{31}\sum_{j=0}^{Nj-1} iS_{i,j}{}^{2}(n) \Big/ \sum_{n=0}^{31}\sum_{j=0}^{Nj-1} S_{i,j}{}^{2}(n) \qquad (7)$$

$$\sigma_g{}^{2}=\frac{1}{Nf}\sum_{i=0}^{Nf-1}(G(i)-Eg)^{2} \qquad (8)$$

$$Eg=\frac{1}{Nf}\sum_{i=0}^{Nf-1}G(i) \qquad (9)$$

$$f_k(\vec{x})=-\frac{1}{2}(\vec{x}-\vec{m}_k)^{T}C_k{}^{-1}(\vec{x}-\vec{m}_k) + [\log p(w_k)-\frac{1}{2}\log|C_k|] \qquad (10)$$

$\vec{x}$ ：入力ベクトル（$\sigma_s{}^{2}$, Eg, $\sigma_g{}^{2}$）
$\vec{m}_k$ ：クラスkの平均値ベクトル
$C_k$ ：クラスkの共分散行列
$p(w_k)$：クラスkの発生確率

【図６】

$$\sigma_e^2 = \frac{1}{NfNj} \sum_{i=0}^{Nf-1} \sum_{j=0}^{Nj-1} (\langle S_{i,j}^2(n) \rangle - \langle S_{i,j}^2(n) \rangle)^2 \qquad (1)$$

$$\langle S_{i,j}^2(n) \rangle = \frac{1}{NfNj} \sum_{i=0}^{Nf-1} \sum_{j=0}^{Nj-1} S_{i,j}^2(n) \qquad (2)$$

Si,j(n)はフレームi，サンプルj，サブバンドnのサブバンドデータ
Nfは単位時間あたりのフレーム数
Njは１フレーム中のサンプル数

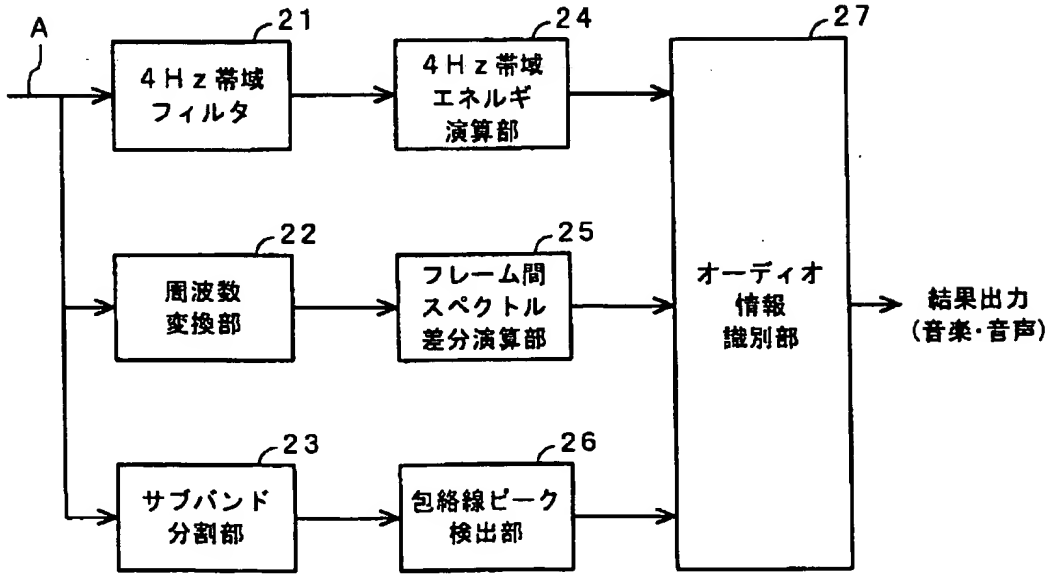
$$SE(i,j) = \sum_{n=0}^{31} S_{i,j}^2(n) \qquad (4)$$

$$NS(i,j) = \begin{cases} 1 & if \quad SE(i,j) > Th1 \\ 0 & else \end{cases} \qquad (5)$$

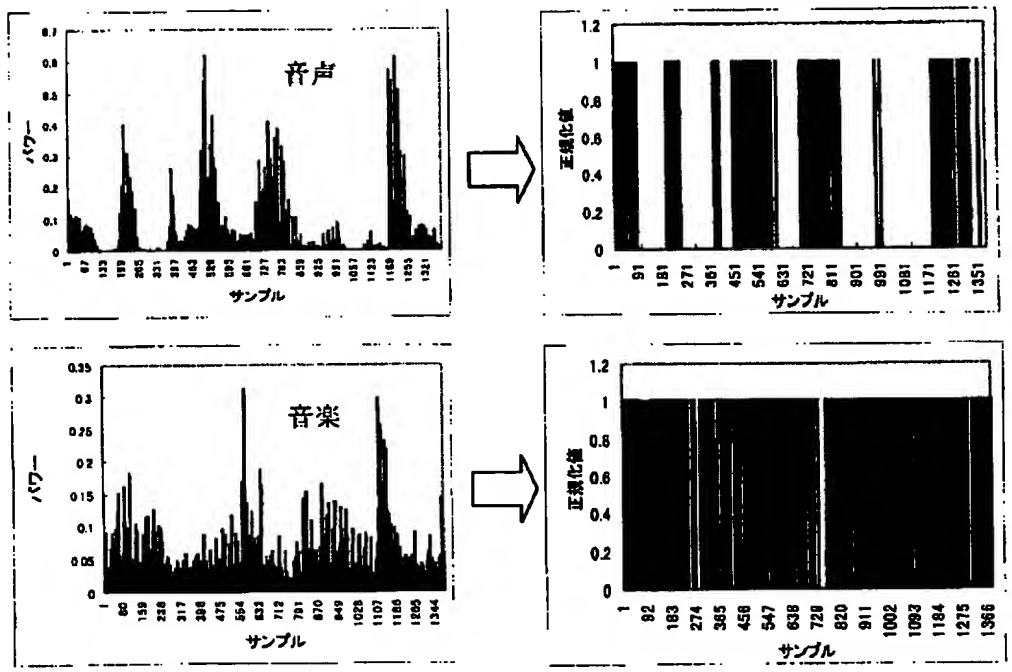
$$\sigma_s^2 = \frac{1}{M} \sum_{k=0}^{M} (N_{ns}(k) - \langle N_{ns}(k) \rangle)^2 \qquad (6)$$

MはＮＳ(i,j)が単位時間内に１から０へ変化する数
$N_{ns}(k)$はＮＳ(i,j)が０となる連続数
$\langle N_{ns}(k) \rangle = \frac{1}{M} \sum_{k=0}^{M} N_{ns}(k)$

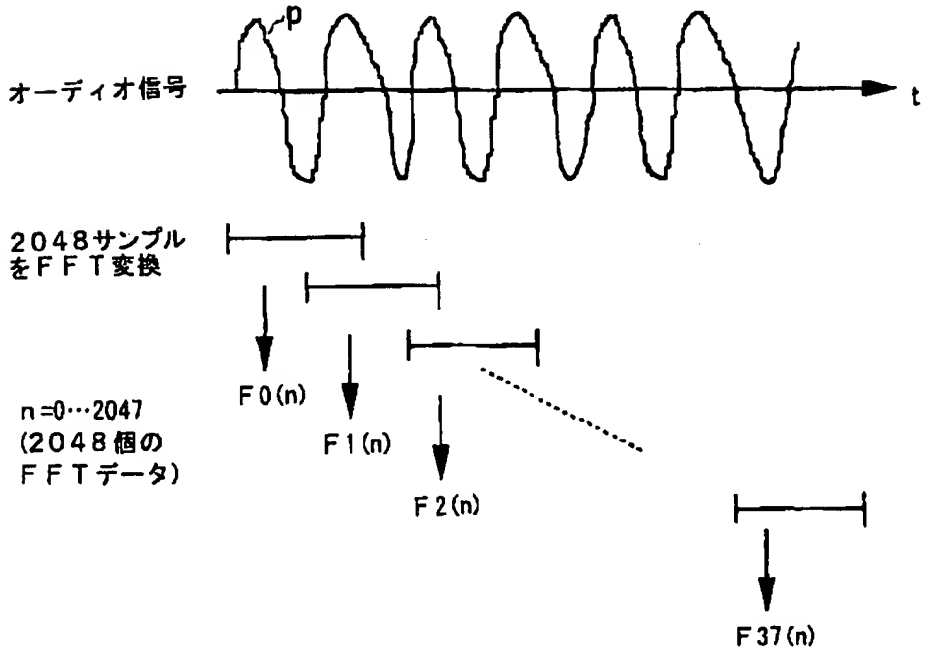
【図１３】



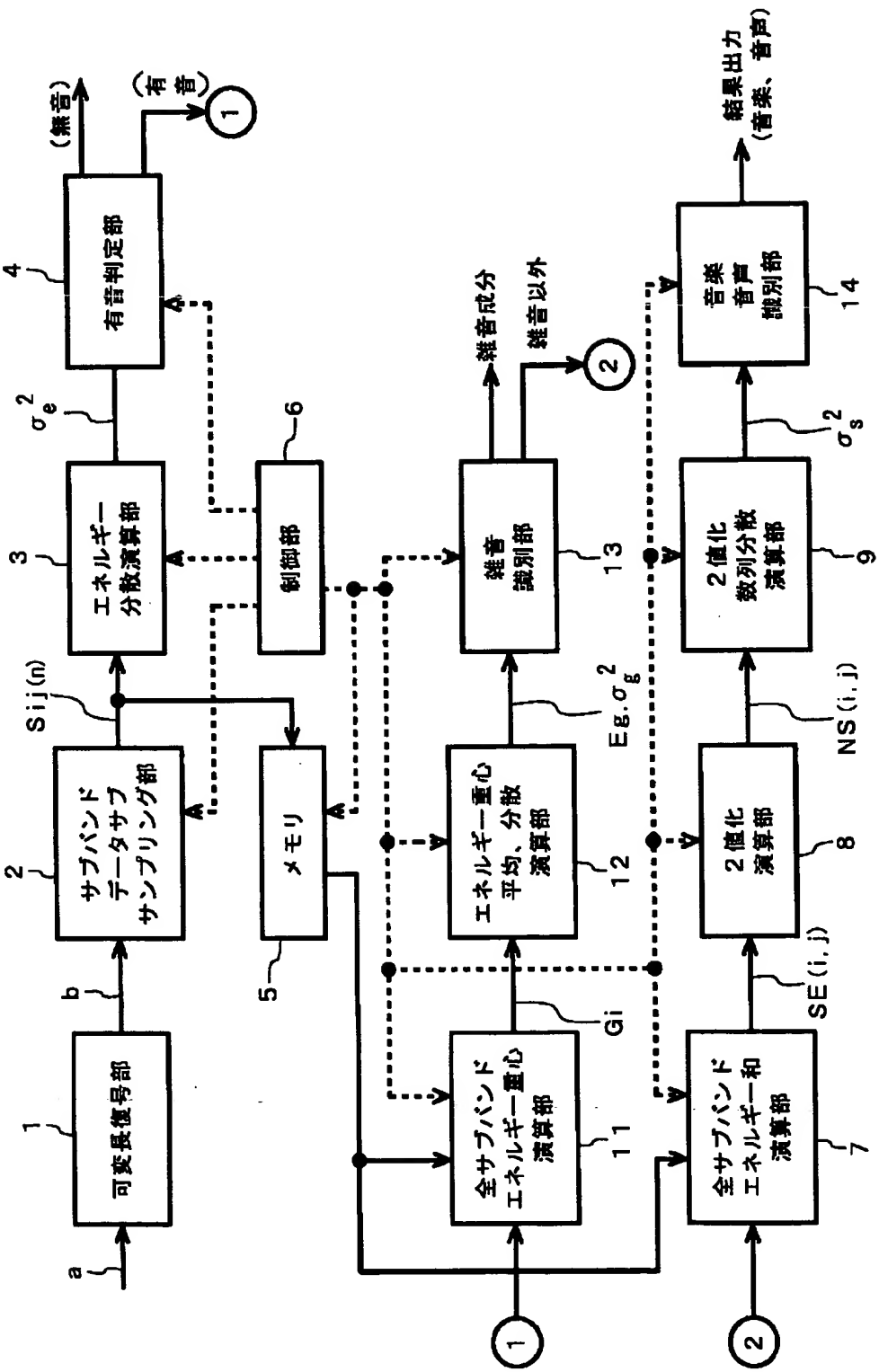
【図８】


音声
音楽
パワー
サンプル
正規化値
正規化前
正規化後

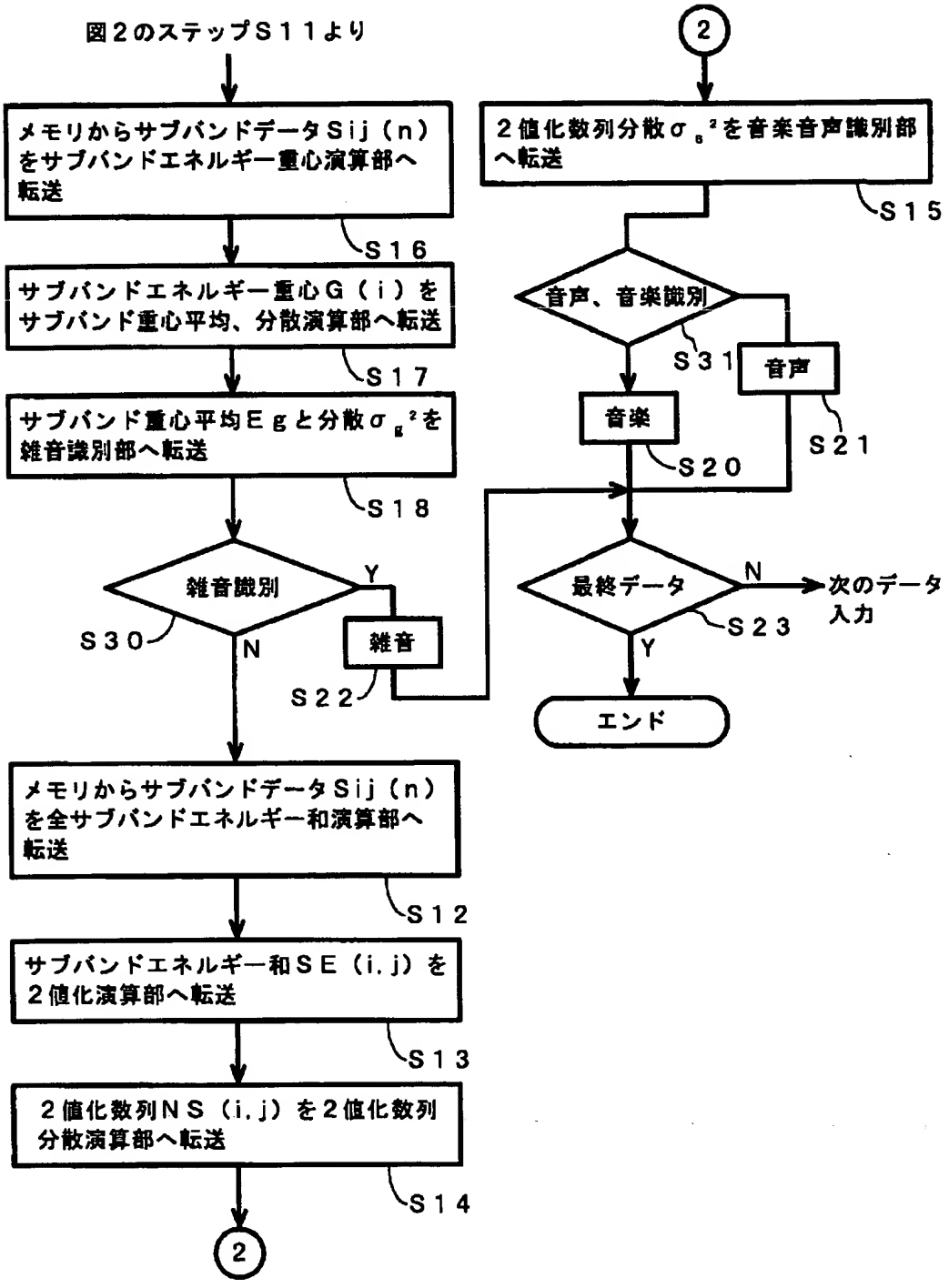
【図１２】


オーディオ信号
p
t
２０４８サンプル
をＦＦＴ変換
n=0…2047
(２０４８個の
ＦＦＴデータ)
F0(n)
F1(n)
F2(n)
F37(n)

【図１０】


可変長復号部
1
a
b
サブバンド
データサブ
サンプリング部
2
エネルギー
分散演算部
3
有音判定部
4
(無音)
(有音)
1
$\sigma_e^2$
Sij(n)
メモリ
5
制御部
6
全サブバンド
エネルギー和
演算部
7
2値化
演算部
8
2値化
数列分散
演算部
9
全サブバンド
エネルギー重心
演算部
11
エネルギー重心
平均、分散
演算部
12
雑音
識別部
13
音楽
音声
識別部
14
雑音成分
雑音以外
2
結果出力
(音楽、音声)
Gi
SE(i,j)
Eg.$\sigma_g^2$
NS(i,j)
$\sigma_s^2$

【図１１】




フロントページの続き

(72)発明者　米山　暁夫
東京都新宿区西新宿２丁目３番２号　国際電信電話株式会社内

(72)発明者　柳原　広昌
東京都新宿区西新宿２丁目３番２号　国際電信電話株式会社内